

Spatial Statistical Software

by

Stephen L. Rathbun
Department of Health Administration, Biostatistics, and
Epidemiology
College of Public Health
University of Georgia
Athens, GA 30605
rathbun@uga.edu

1

General Observations:

- Perhaps the single most limiting factor for dissemination a modern spatial statistical procedures is the limited availability of statistical software.
- Writing of statistical software involves the following trade-off:
 - Ease of use
 - Flexibility

2

Outline: Review software for three areas of spatial statistics.

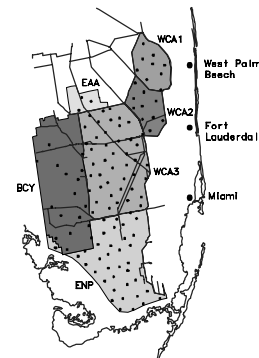
1. Geostatistics.
2. Spatial Point Patterns.
3. Lattice Data.

3

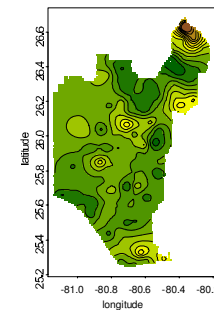
Geostatistics

South Florida Ecosystem Assessment.

Sample Sites



Predicted Total Mercury



4

- SAS
- Surfer
- ArcGIS Geostatistical Analyst
- S+SpatialStats
- R

% SAS

File Edit View Tools Solutions Window Help

Results

Results

- Qplot: The SAS System
- Qplot: The SAS System
- Qplot: The SAS System
- Qplot: Variogram TOC Data
- Macro: Variogram TOC Data
- Qplot: Variogram TOC Data

Output - (Unlinked)

Variogram TOC Data

The NLIN Procedure

NOTE: An intercept was not specified

Source	DF	Sum of Squares	S
Model	2	4926.0	2
Error	16	78.5505	4
Uncorrected Total	18	684.6	

Parameter	Estimate	Approx Std Error
ce	57.2718	2.4360
alpha	11.0140	1.8924

Macro % PROC GMLT (running)

```

title "Variogram TOC Data";
data toc;
  infile "c:\geo3\ch04\wago\soe.txt";
  input x y depth toc;
proc variogram outv=var;
  compute lagd=4.9 maxlags=18;
  coordinates xc=x yc=y;
  var toc;
proc nlin method=dlad;
  parameters ce=60 alpha=9;
  bounds ce>0, alpha>0;
  tempce=(1-exp(-distance/alpha));
  model variag=temp;
  weight _weight/temp;
  output outb=varoutid p=varb;
run;
proc gplot;
  axial order=0 to 80 by 10 length=4 in;
  axial order=0 to 90 by 10 length=4 in;
  plot (variag vhat)*distance / overlay vaxis=axis1 haxis=axis2;
  symbol1 v=J special h=2 i=none color=black;
  symbol2 v=none i=1 l=spline color=black;
run;

```

GRAPH1 WORK.G001.GPLOT1

Variogram TOC Data

Output - (Unlinked)

Log - (Unlinked)

GRAPH1 WORK.G001.GPLOT1

Log3 SAS % PROC GMLT

6

The screenshot displays the SAS Enterprise Guide interface. On the left, a project tree shows a dataset named 'Results'. The main window, titled 'Output - (Untitled)', shows the output of a REML estimation procedure. The output is organized into several sections: 'REML Estimation of Variogram Parameter TOC Data', 'The Mixed Procedure', 'Covariance Parameter Estimates', 'Fit Statistics', 'PWRMS Model Likelihood Ratio Test', and 'Solution for Fixed Effects'.

REML Estimation of Variogram Parameter TOC Data
09:35 Friday, February 3, 2006

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SP(COP)	Intercept	19.5293
Residual		62.0973

Fit Statistics

Statistic	Value
-2 Res Log Likelihood	688.5
AIC (smaller is better)	692.5
AICC (smaller is better)	692.6
BIC (smaller is better)	697.8

PWRMS Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	6.77	0.0093

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	24.9400	1.9531	0	12.77	.

The bottom of the window shows the SAS code used to generate the output:

```

title "REML Estimation of Variogram Parameters TOC Data";
data too;
  infile "c:\projects\wags\too.txt";
  input x y depth too;
run;

proc mixed;
  model too=solution;
  parms (57.3) (11.0);
  repeated / subject=intercept type=exp (x y);
run;
  
```

The screenshot displays the SAS Studio interface. The top menu bar includes File, Edit, View, Tools, Run, Solutions, and Window. The top toolbar contains icons for file operations, editing, and running. The main window is divided into three panes:

- Results Pane (Top):** Shows the output of the PROC KRIG2D statement. The title is "Kriging TOC Data" with a timestamp of "09:36 Friday, February 3, 2006". The output is a table with 10 columns: Obs, LABEL, VARNAME, GXC, GYC, NPOINTS, ESTIMATE, and STDERR. It lists 29 observations for the variable TOC, showing predicted values and standard errors for various locations (e.g., Pred1.Model11, Prc1.Model11, etc.).
- Editor Pane (Bottom):** Contains the SAS script used to generate the results. The script includes a title statement, a data step to read a file, a PROC KRIG2D statement with various options (outest=tochat, predict, var=toc, model, gcid, covlikelihood, sum=p), and a run statement.
- Log Pane (Bottom):** Shows the SAS log output, which includes the same PROC KRIG2D statement and the log of the model fitting process.

The PROC KRIG2D statement in the script is as follows:

```
proc krige2d outest=tocchat;
  predict var=toc;
  model nugget=0 scale=62.8973 range=10.9939 covexpexp;
  gcid x=1690 to 1725 by 25 y=-1675 to -1500 by 25;
  covlikelihood sum=p;
run;
```

195

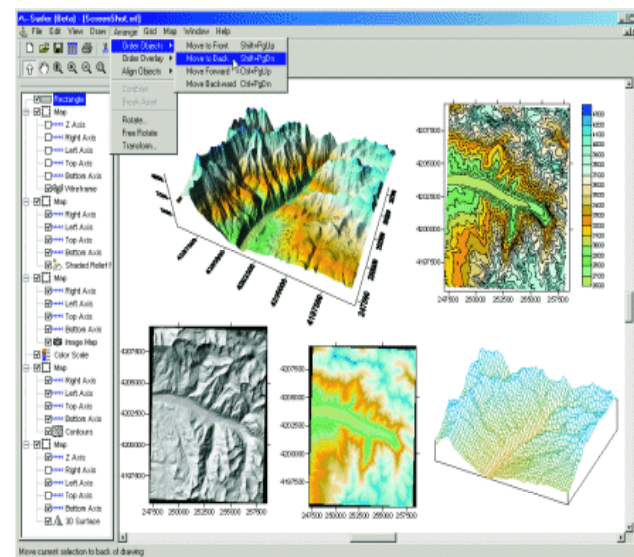
Comments: SAS Geostatistics

- SAS is not menu driven. Analysis is carried out by writing SAS programs in the SAS editor.
 - For those who have experience with SAS, the geostatistical procedures are easy to apply.
 - Harder to use than menu-driven software.
- SAS has procedures for:
 - Isotropic and anisotropic variogram estimation (proc variog);
 - Variogram model fitting:
 - ▶ Weighted Least Squares (proc nlin);
 - ▶ Maximum Likelihood and REML (proc mixed).
 - Ordinary Kriging (proc krige2d).
 - Universal Kriging (proc mixed).
 - Generalized Mixed Models (proc glimmix).
- Limitations:
 - Limited choice of variogram models.
 - Cannot draw good contour maps.

9

Surfer

<http://www.goldensoftware.com/>



10

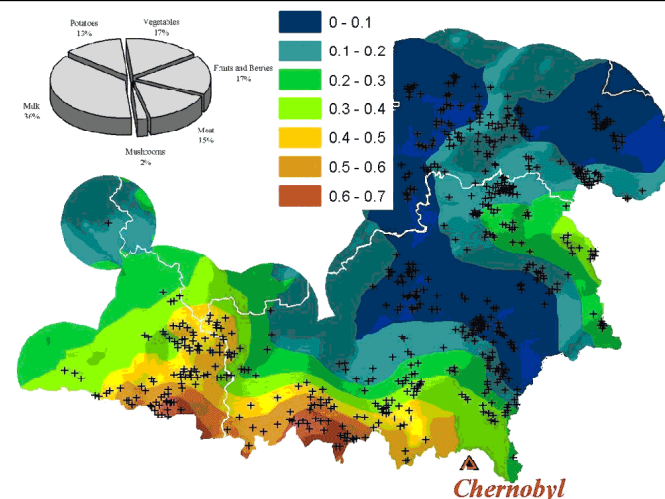
Comments: Surfer

- Menu Driven
- Surfer has procedures for:
 - Variogram Estimation;
 - Least Squares Estimation of Variogram Model Parameters
 - Wide Variety of Variogram Models: exponential, Gaussian, linear, log, power, quadratic, rational quadratic, spherical, wave, pentaspherical, cubic.
 - Ordinary Kriging
 - Excellent mapping capabilities: contour maps; 3D surface maps; wireframe maps; vector maps; shaded relief maps.
- Limitations:
 - Cannot fit Matern variogram;
 - Universal kriging not available.

11

ArcGIS Geostatistical Analyst

<http://www.esri.com/software/arcgis/extensions/geostatistical/index>



12

Comments: ArcGIS Geostatistical Analyst

- Menu Driven
- Geostatistical Analyst has procedures for:
 - Isotropic and anisotropic variogram estimation;
 - Least squares estimation of variogram parameters;
 - Wide variety of variogram models: circular, spherical, tetraspherical, pentaspherical, exponential, Gaussian, rational quadratic, hole effect, k-bessel, stable.
 - Variety of kriging methods:
 - ▶ Ordinary kriging
 - ▶ Universal kriging
 - ▶ Indicator kriging (Binary Variables)
 - ▶ Disjunctive kriging (Nonlinear Geostatistics)
 - ▶ Cokriging (Multivariate Geostatistics)
 - Crossvalidation for model diagnostics.
- Limitation: Expensive (\$2,500 for Geostatistical Analyst, \$1,500 for ArcView 9.1)

13

Definition: Crossvalidation.

- Remove the data at site s_i from the data set;
 - Use the remaining data to obtain the kriging predictor $\hat{Z}_{-i}(s_i)$ of the data at site s_i
 - Compute the corresponding kriging variance $\sigma_{-i}^2(s_i)$
- Repeat the above procedure for all sites.
- Compare observed values $Z(s_i)$ with predicted values $\hat{Z}_{-i}(s_i)$
 - Bias Measure

$$CV_1 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{Z(s_i) - \hat{Z}_{-i}(s_i)}{\sigma_{-i}(s_i)} \right\}$$

- Uncertainty Assessment

$$CV_2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{Z(s_i) - \hat{Z}_{-i}(s_i)}{\sigma_{-i}(s_i)} \right\}^2$$

For a valid model, we should have

$$CV_1 \cong 0 \text{ and } CV_2 \cong 1$$

14

S+SPATIALSTATS

<http://www.insightful.com/products/spatial/default.asp>

Variogram Estimation

Least Squares Estimation

15

S+SPATIALSTATS

Ordinary Kriging

Universal Kriging

16

Comments: S+SPATIALSTATS

- Menu Driven
- S+SPATIALSTATS has procedures for:
 - Isotropic and Anisotropic Variogram Estimation
 - Least Squares Estimation of Variogram Parameters (Weighted least squares with some work)
 - Limited variogram models: Spherical, exponential, Gaussian
 - Ordinary and Universal Kriging
 - Good quality contour maps
- Software has not been kept up to date.
 - Effort has been made to improve user interface.
 - No effort has been made to include modern methods.

17

R

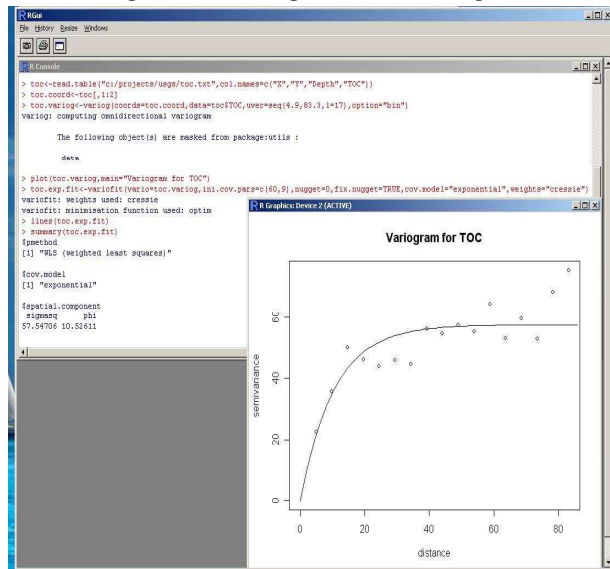
<http://www.r-project.org/>

Geostatistical Packages

- geoR <http://www.est.ufpr.br/geoR/>
Frequentist and Bayesian geostatistics.
- geoRglm <http://www.daimi.au.dk/~olefc/geoRglm/>
Geostatistics for counts data. Poisson and binomial models.
- fields <http://www.image.ucar.edu/GSP/Software/Fields/>
Best for global data. Includes great circle distance.
- gstat <http://www.gstat.org/>
- RandomFields
http://www2.hsu-hh.de/schlath/R/RandomFields/RandomFields_doc.
Spatial simulation.

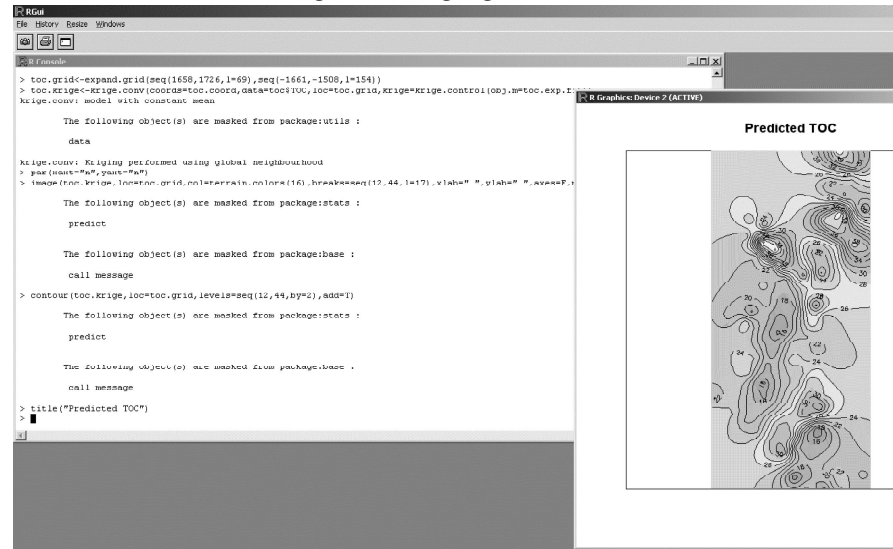
18

geoR: Variogram Modeling



19

geoR: Kriging



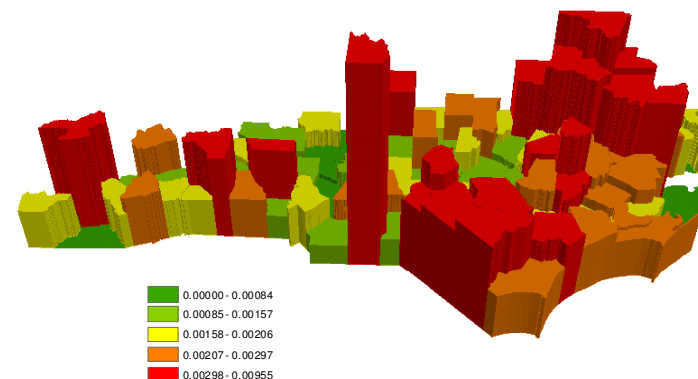
20

Comments: R

- Public domain software;
- Packages contributed by statistical researchers keep the software up to date;
- Command driven and interactive;
- GeoR has procedures for:
 - Variogram estimation;
 - Least squares, weighted least squares, REML estimation of variogram parameters;
 - Bayesian inference for model parameters;
 - Diverse variety of variogram models including the Matérn class;
 - Ordinary, universal and Bayesian kriging.
- GeoRglm has procedures for binomial and Poisson models for counts data;
- Fields includes great circle distance for investigating global data;
- Limitation: Not well documented.

21

Lattice Data Sudden Infant Death Rates in North Carolina



22

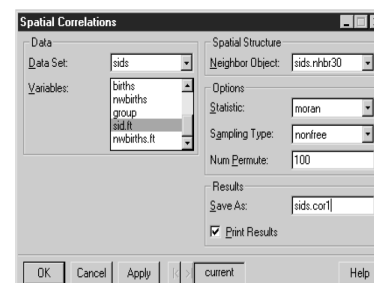
Lattice Model Software

- S+SPATIALSTATS
- BUGS
- R package: spdep

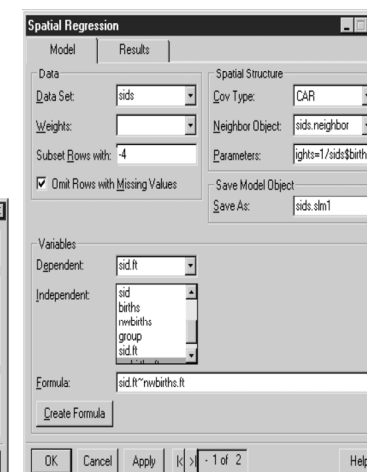
23

S+SPATIALSTATS Lattice Models

Moran's Index



CAR Model



24

Comments: S+SPATIALSTATS

- Menu Driven
- S+SPATIALSTATS has procedures for:
 - Defining neighborhood matrices
 - Defining spatial weights matrices
 - Computing Moran's I
 - Fitting spatial regression models:
 - ▶ Conditional AutoRegressive
 - ▶ Simultaneous AutoRegressive
 - ▶ Moving Average

25

GeoBUGS

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs.shtml>



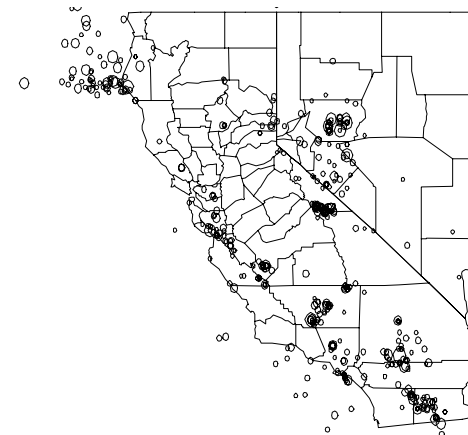
26

Comments: GeoBUGS

- Public domain software;
- Bayesian inference for lattice models:
 - CAR models
 - Poisson and binomial models with spatially dependent random effects.
- Data interface can use some work.

27

Spatial Point Pattern California Earthquakes



28

Point Pattern Software:

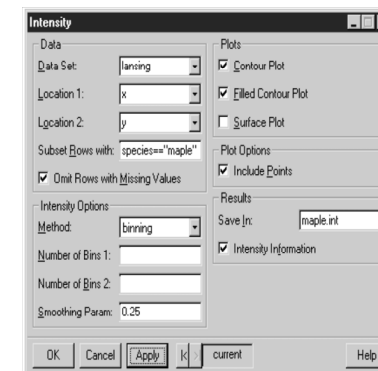
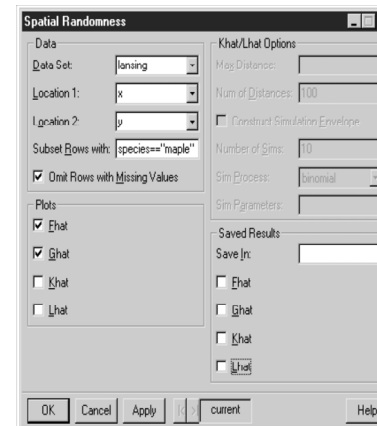
- S+SPATIALSTATS
- R

29

S+SPATIALSTATS

K-Function

Nonparametric Intensity



30

Comments: S+SPATIALSTATS

- Menu Driven;
- S+SPATIALSTATS has procedures for:
 - Computing F-, G- and K-functions;
 - Testing complete spatial randomness;
 - Nonparametric estimation of the intensity function;
 - Fitting the point cluster process model.

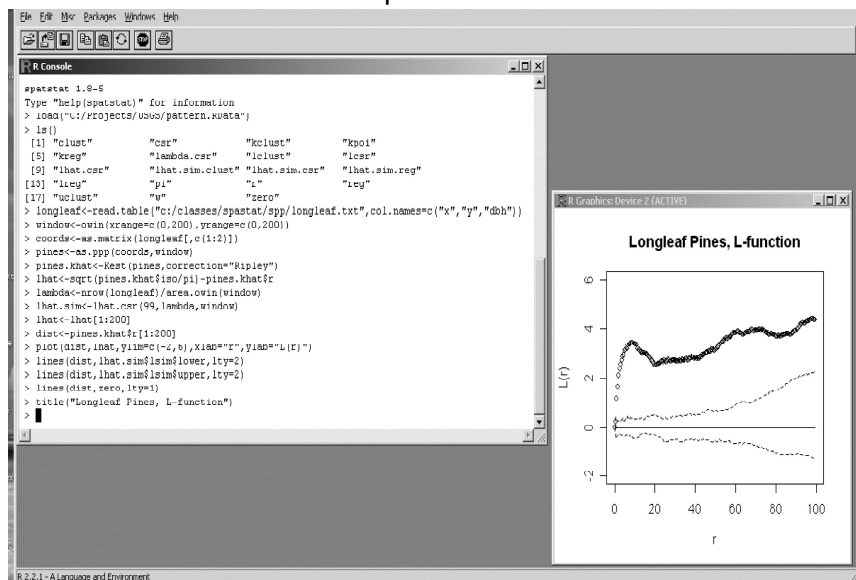
31

R: Point Pattern Packages:

- spatstat <http://www.spatstat.org/>
Analysis of spatial point patterns.
- splancs <http://www.maths.lancs.ac.uk/~rowlings/Splancs/>
Analysis of spatial and spatiotemporal point patterns.
- MarkedPointProcess
<http://www2.hsu-hh.de/schlath/schlather.html#Software>
Analysis of marked point patterns.

32

R: spatstat



33

Comments: R

- Public domain software;
- Packages contributed by statistical researchers keep the software up to date;
- Command driven and interactive;
- Spstat has procedures for:
 - Computing F-, G- and K-functions;
 - Testing complete spatial randomness;
 - Fitting the point cluster process model;
 - Simulating a variety of point process models;
 - Estimating parameters of modulated Poisson process model (covariates must be observed at all locations).

34

General Summary

- ArcGIS Geostatistical Analyst:
 - Menu driven;
 - A comprehensive collection of geostatistical methods;
 - Expensive.
- R:
 - Up-to-date methods for geostatistical and point pattern analyses;
 - Public domain;
 - Command driven and interactive.
- S+SPATIALSTATS:
 - Best for analysis of lattice data;
 - Menu driven.

35